



LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

Two-Sided Pyramid Wavefront Sensor in the Direct Phase Mode

D. Phillion, K. Baker

April 20, 2006

Astronomical Telescopes and Instrumentation 2006 SPIE
Meeting
Orlando, FL, United States
May 24, 2006 through May 31, 2006

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

[†]Two-Sided Pyramid Wavefront Sensor in the Direct Phase Mode

Donald W. Phillion and Kevin Baker
Lawrence Livermore National Laboratory

ABSTRACT

The two-sided pyramid wavefront sensor has been extensively simulated in the direct phase mode using a wave optics code. The two-sided pyramid divides the focal plane so that each half of the core only interferes with the speckles in its half of the focal plane. A relayed image of the pupil plane is formed at the CCD camera for each half. Antipodal speckle pairs are separated so that a pure phase variation causes amplitude variations in the two images. The phase is reconstructed from the difference of the two amplitudes by transforming cosine waves into sine waves using the Hilbert transform. There are also other corrections which have to be applied in Fourier space. The two-sided pyramid wavefront sensor performs extremely well: After two or three iterations, the phase error varies purely in y . The two-sided pyramid pair enables the phase to be completely reconstructed. Its performance has been modeled closed loop with atmospheric turbulence and wind. Both photon noise and read noise were included. The three-sided and four-sided pyramid wavefront sensors have also been studied in direct phase mode. Neither performs nearly as well as does the two-sided pyramid wavefront sensor.

Keywords: adaptive optics, pyramid wavefront sensor

1. INTRODUCTION

The pyramid wave-front sensor was only recently proposed¹ and its operation as a direct phase measuring device only very recently realized^{2,3}. Pyramid sensors inherit properties of both slope sensors and direct phase sensors. They can in principle be used in slope mode to bootstrap to a high Strehl ratio regime and then operate in a direct phase sensing mode, which has a much lower wavefront reconstruction error for low photon fluxes. The pyramid wave-front sensor has been compared in both regimes to the spatially filtered Shack-Hartmann wavefront sensor⁴. The pyramid wave-front sensor in the direct phase mode acts similarly in many ways to a white-light interferometer, but even though its operation in this mode is interferometric in nature, it acts fundamentally different from familiar interferometers such as the Mach-Zehnder interferometer. In the direct phase mode, what it measures is not the phase itself, but rather the Hilbert transform of the phase. Since the interference is close to zero-path-length-difference, it is appropriate for the temporally incoherent light encountered in astronomical applications.

The pyramid wavefront sensor is placed at the focal plane with the guide star focused on the apex for an n -sided pyramid wavefront sensor with $n \geq 3$ or on the ridge for a 2-sided pyramid wavefront sensor. Each speckle in the point-spread-function corresponds to a plane wave component at the telescope pupil. A small sinusoidal phase fluctuation or a small sinusoidal amplitude fluctuation at the telescope pupil gives rise to an antipodal pair of speckles having the same magnitude. If the sinusoidal fluctuation is one in both amplitude and phase, then the speckles are of unequal magnitude. The speckle pair partners are split apart by the pyramid wavefront sensor and each speckle interferes with a piece of the core. For each pyramid wavefront sensor sector, the telescope aperture is imaged onto a CCD so that the CCD is at a plane conjugate to the telescope aperture. Each speckle causes a periodic amplitude variation. For a sinusoidal phase fluctuation, the sinusoidal amplitude fluctuations in each sector are ninety degrees out of phase with the original sinusoidal phase fluctuation. This represents a Hilbert transform. Actually, things are more complicated than this because each piece of the core has a different centroid than the original core and this causes shifts for all the spatial frequencies. Fortunately, this can be corrected in the reconstruction for the 2-sided pyramid wavefront sensor.

[†] This document UCRL-PROC-220733 was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

As we shall show later, we found for the 2-sided pyramid WFS that the reconstruction obtained from the amplitude difference of the two sides was greatly more accurate than the reconstruction obtained from the intensity difference of the two sides.

The CCD camera measures intensity not amplitude, however. Amplitude is just the square root of intensity except that there is the ambiguity of which sign to use. The Strehl ratio must be high enough so that the amplitude does not go through zero and reverse sign. In other words, in each sector at the CCD the field due to the speckles must never exceed in magnitude the field due to the core.

Another important limitation that any interferometric wave sensor has is that branch errors are possible. If one mems actuator is displaced by $\lambda/2$ from where it is supposed to be, causing a λ change in the reflected wavefront, this error will persist. Even for the extreme AO systems being considered, it will not be possible to phase unwrap the accumulated mems wavefront changes. Most pixels will have wavefront phase jumps of more than $\lambda/2$ to at least one of their neighbors. Only if there are two interferometric wavefront sensors operating at very different wavelengths will it be possible to phase unwrap the wavefront and discover such errors. Fortunately, having a pixel get onto the wrong branch is extremely unlikely when the frame rate is high and the guide star is bright.

Classic pyramid sensors operate with a four-sided pyramid producing four separate images. In the direct phase mode, wavefront reconstruction is more accurate if the sensor produces only two separate images, splitting the light with a ridge rather than a pyramid. This is the type of wavefront sensor modeled in this paper, but we will continue to refer to it as a “pyramid” rather than a “ridge” wavefront sensor.

2. PRINCIPLES OF OPERATION

The equations describing the interferometric mode for pyramid sensing are described below. Small phase and amplitude fluctuations in the pupil may be represented by

$$\begin{aligned} z^A &= 1 + \sum_n 0.5\epsilon_n^A \left[\exp(i(\vec{k}_n \cdot \vec{x} + \theta_n^A)) - \exp(-i(\vec{k}_n \cdot \vec{x} + \theta_n^A)) \right] \\ &= 1 + i \sum_n \epsilon_n^A \sin(\vec{k}_n \cdot \vec{x} + \theta_n^A) \text{ for phase} \end{aligned} \quad (1)$$

$$\begin{aligned} z^B &= 1 + \sum_n 0.5\epsilon_n^B \left[\exp(i(\vec{k}_n \cdot \vec{x} + \theta_n^B)) + \exp(-i(\vec{k}_n \cdot \vec{x} + \theta_n^B)) \right] \\ &= 1 + \sum_n \epsilon_n^B \cos(\vec{k}_n \cdot \vec{x} + \theta_n^B) \text{ for amplitude} \end{aligned} \quad (2)$$

Each focal plane speckle pair is represented by a \mathbf{k}_n vector and a $-\mathbf{k}_n$ vector. The \mathbf{k}_n vector is chosen to be in the right half side of the focal plane, so that $k_x > 0$. Speckle pairs on the k_y axis can't be measured so we ignore this possibility. The epsilon quantities are to be regarded as infinitesimals. After passing through a two-sided pyramid with a vertical ridge, the left-half-side (LH) and right-half-side (RH) amplitudes will be:

$$\begin{aligned} \text{LH} &= 1 - \sum_n 0.5\epsilon_n^A \cos(\vec{k}_n \cdot \vec{x} + \theta_n^A) \text{ for phase} \\ \text{RH} &= 1 + \sum_n 0.5\epsilon_n^A \cos(\vec{k}_n \cdot \vec{x} + \theta_n^A) \text{ for phase} \end{aligned} \quad (3)$$

$$\begin{aligned} \text{LH} &= 1 + \sum_n 0.5\epsilon_n^B \cos(\vec{k}_n \cdot \vec{x} + \theta_n^B) \text{ for amplitude} \\ \text{RH} &= 1 + \sum_n 0.5\epsilon_n^B \cos(\vec{k}_n \cdot \vec{x} + \theta_n^B) \text{ for amplitude} \end{aligned} \quad (4)$$

These equations are overly simplistic and miss important parts of the physics. The core in the focal plane is not a point, but rather is an Airy pattern. The two halves of this Airy pattern both have different centroids than that of the whole Airy pattern. This leads to shifts in the k_x components of the wave vectors. Also, the diffraction pattern of half an Airy disk is not a uniform

spherical wave. Neither is it a top hat at a plane conjugate to the telescope aperture plane. The reference wave generated by each half Airy pattern asymmetrically brightens near the edge and extends asymmetrically far beyond the geometric image of the telescope aperture. Nevertheless, we will work with these equations.

The difference of the left-half-side (LH) and right-half-side (RH) amplitudes provides the phase fluctuations,

$$RH - LH = \begin{cases} \sum_n \epsilon_n^A \cos(\vec{k}_n \cdot \vec{x} + \theta_n^A) & \text{for phase} \\ 0 & \text{for amplitude} \end{cases} \quad (5)$$

and the sum of the left-half-side (LH) and right-half-side (RH) provides the amplitude fluctuations

$$RH + LH = \begin{cases} 2 & \text{for phase} \\ 2 + \sum_n \epsilon_n^B \cos(\vec{k}_n \cdot \vec{x} + \theta_n^B) & \text{for amplitude} \end{cases} \quad (6)$$

The term represented by the constant “2” in Eq (6) represents the diffraction pattern of the half core reference wave at the CCD plane. Remember that the CCD plane is conjugate to the telescope aperture plane. This term is actually nonuniform with asymmetric edge brightening. This makes extracting the low order amplitude variations difficult. However, the core diffraction pattern subtracts out in the difference represented by Eq (5). The reference wave is nonuniform in both amplitude and phase. The part of the phase nonuniformity that is due to the centroid shift causes a constant shift in all the k_x values.

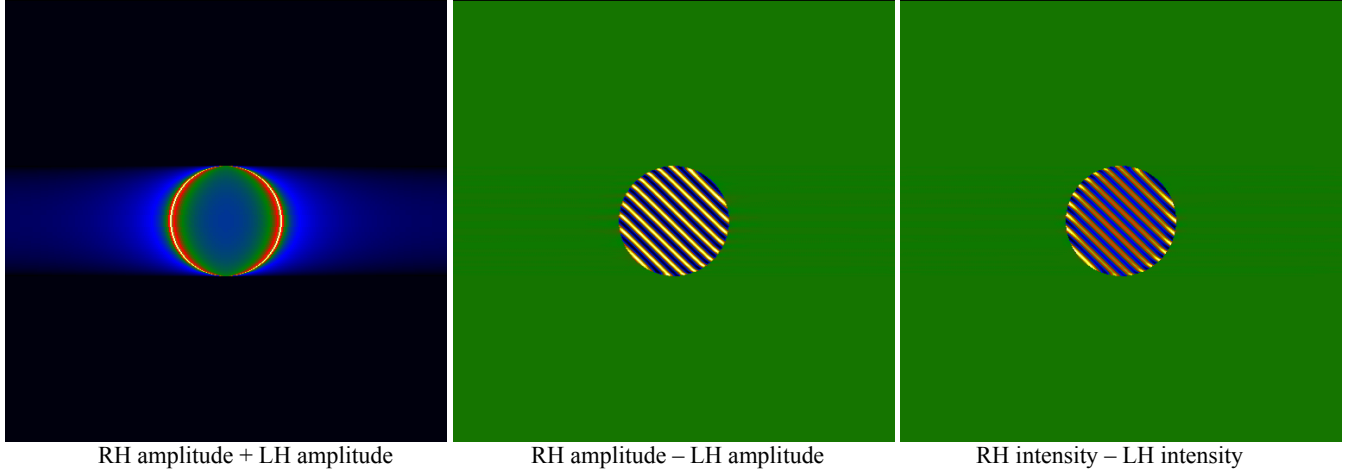


FIGURE 1 LH and RH obtained from the sinusoidal phase fluctuation $0.1 \cos(x+y)$ radians with $D=50$ $W=200$ 512×512

Figure 1 shows the RH+LH amplitude sum, the RH-LH amplitude difference, and the RH-LH intensity difference for the sinusoidal phase fluctuation $0.1 \cos(x+y)$ radians in a circular pupil with diameter $D=50$ mm on a 200×200 mm, 512×512 pixel grid. The RH-LH amplitude difference and the RH-LH intensity difference are shown side-by-side in order to demonstrate how much better the reconstruction is when one reconstructs from the amplitude difference rather than the intensity difference. The explanation comes from the fact that the reference wave is the diffraction pattern of a half Airy pattern and so is not constant. If one simply takes the amplitude difference, the reference wave subtracts out, whereas if one takes the intensity difference, then the reference wave amplitude is a factor in the interference term.

For this example, the amplitude was one within the circular pupil and zero outside. The focal plane core was approximately 4 pixels across. The LH and RH complex wavefronts were generated by Fourier transforming the phase screen complex wavefront, taking the left half side and inverse Fourier transforming to obtain the LH complex wavefront, and similarly for the RH complex wavefront. Pixels on the k_y axis in Fourier space were divided equally between the LH and RH.

The sinusoidal amplitude fluctuations in RH-LH given by Eq (5) are ninety degrees out of phase with the original sinusoidal phase fluctuations given by Eq (1). Reconstruction of the phase requires that cosine waves be transformed into sine waves. In Fourier space, Fourier plane wave components with $k_x > 0$ are multiplied by $+i$ and Fourier plane wave components with $k_x < 0$ are multiplied by $-i$. This is the Hilbert transform. Fourier components with $k_x = 0$ cannot be reconstructed.

There is the additional complication that in the focal plane the centroids of the core halves are displaced from the centroid of the whole core. This necessitates correcting the k_x components of the spatial frequencies for the real Fourier transform of the RH-LH amplitude difference. When adjusting the frequency of a cosine or sine wave component, we want its phase value at the center of the pupil to remain unchanged. Here is what we do to accomplish this: First the RH-LH amplitude difference image is translated so that the center of the image is at the (0,0) pixel at the top left corner. The image is assumed to be periodic in x and y with the grid size. Then the two-dimensional real Fast Fourier Transform is performed. The Fourier transformed image is then translated so that the pixel at (0,0) is now at $(n_x/2, n_y/2)$ where n_x is the number of columns and n_y is the number of rows. Pixels in the left half plane are multiplied by $-i$ and pixels in the right half plane are multiplied by $+i$. Pixels on the k_y axis with $k_x=0$ are multiplied by 0. Then pixels on the left side are shifted further to the left by one pixel and pixels on the right side are shifted further to the right by one pixel. Zero values are shifted in from the k_y axis and values shifted off the left and right edges are lost. One pixel shifts give the correct k_x shifts when $D/W=1/4$, where W is the grid size. The reason that is possible to do the spatial frequency corrections at this point is that the core is subtracted out in the LH-RH difference. Then the Fourier transformed image is translated so that the zero frequency pixel is at the (0,0) pixel in the top-left corner. Then the inverse Fast Fourier Transform is performed and the real part is taken. Now cosine wave fluctuations in the RH-LH amplitude difference have become sine wave fluctuations. Also, the k_x components of these sine wave components have been shifted to same signed but larger magnitude k_x values in such a way that the phases at the pupil center have not changed. A sinusoidal fractional amplitude fluctuation now represents a sinusoidal phase fluctuation in radians. The reconstruction is complete. Figure 2 is a flow chart of these computations.

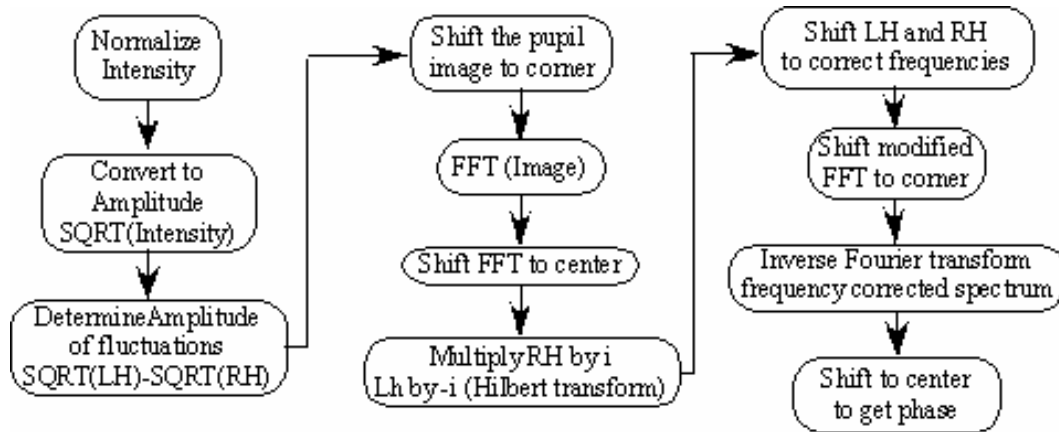


FIGURE 2 Flowchart of the computations for the pyramid wavefront sensor's interferometric reconstruction of the phase.

We have not discussed the first step in this flow chart, which is to normalize the LH and RH CCD intensities. In the simulations we present later, the amplitude in the circular pupil always has average value one, but this of course won't be the case in actual operation. The CCD cameras will have to be flat fielded and registered with respect to each other. Then the intensities will be normalized by that value necessary to make the average intensity RH+LH in the circular pupil be unity. Then the square roots of the intensities will be computed and the RH-LH amplitude difference computed.

Performing the k_x corrections is critical to obtaining good performance with the 2-sided pyramid wavefront sensor. Figure 3 shows the reconstruction errors with and without the k_x corrections while figure 4 compares centered horizontal lineouts for the original phase screen to the reconstruction with and without the k_x corrections. This figure dramatically proves that the k_x spatial frequency values are only reconstructed correctly if the k_x corrections are applied. These reconstruction errors are for one iteration. It might be thought that the errors due to not correcting would be rapidly reduced with further iteration but this is not the case, as Table 1 proves.

	No k_x corrections	With k_x corrections
1	0.0351	0.0102
2	0.0248	0.0032
3	0.0194	0.0012
4	0.0162	0.0005

Table 1

Comparison of the convergence for the first 4 iterations with and without the k_x corrections.

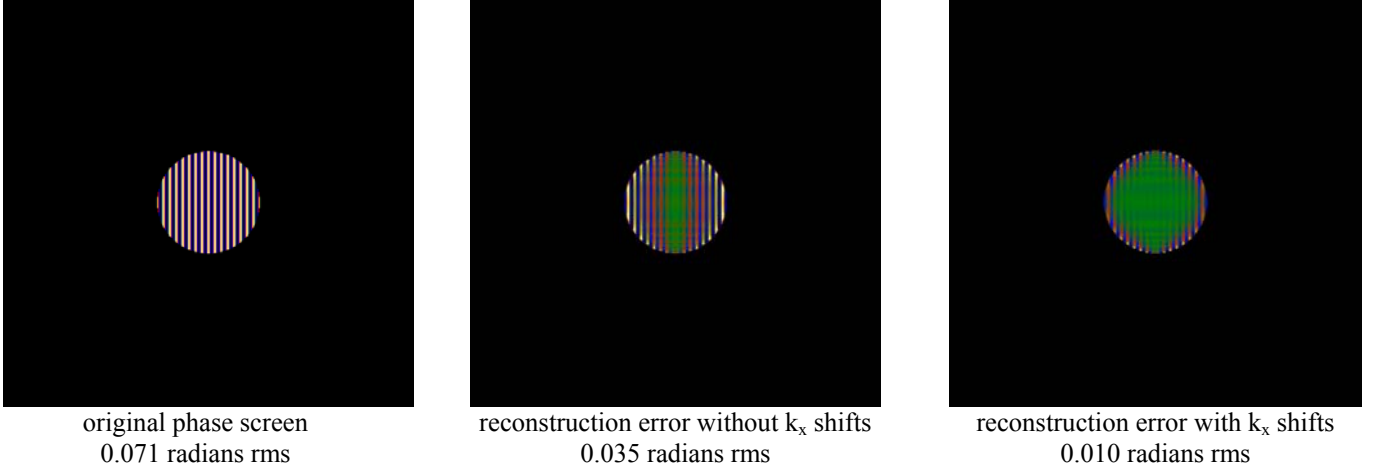


FIGURE 3 Comparison of the reconstruction errors with and without the k_x shift corrections. The phase screen is $0.1 \cdot \cos(2 \cdot x)$ with $D=50$, $W=200$, 512×512

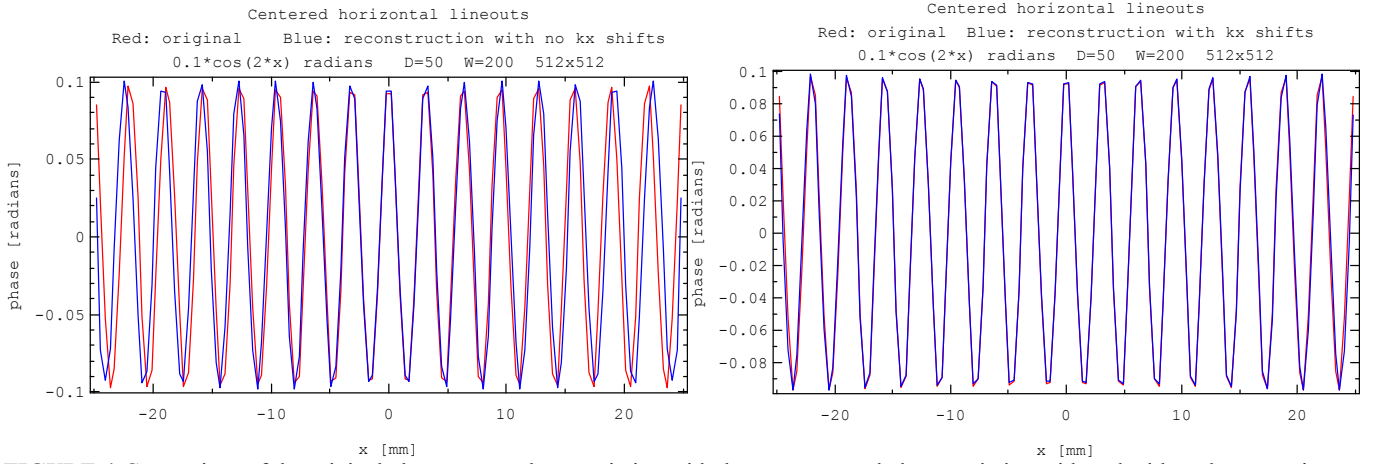


FIGURE 4 Comparison of the original phase screen phase variation with the reconstructed phase variation with and without k_x correction.

3. NOISE PROPERTIES

We first derive the expected rms wavefront reconstruction error due to photon and read noise and then compare it with simulations. For a 2-sided pyramid wavefront sensor with a vertical ridge, the core is split vertically into two parts. Not only does each half of the core get only half the photons that are in the whole core, but because each piece of the core has only half the width, the light diffracts horizontally to about twice the pupil diameter at the CCD. Therefore we expect the amplitude to only be about $\frac{1}{2}$ of what it would be if the core were not divided. In fact, of course, things aren't this simple because the diffraction pattern is complicated with asymmetric brightening near the imaged aperture edge at the CCD, but this is a workable approximation. Let us assume that the phase screen has constant phase and amplitude. Then the only fluctuations are due to the photon and read noise. We will do the photon noise only case first. The generalization to include read noise is then straightforward. Let the phase screen have a single sinusoidal phase perturbation of ε peak radians:

$$1 + i\varepsilon \sin(kx + \theta) \quad (6)$$

Then the LH and RH amplitudes at the CCD are:

$$\begin{aligned} \text{LH} &= \frac{1}{2} - \frac{\varepsilon}{2} \cos(kx + \theta) \\ \text{RH} &= \frac{1}{2} + \frac{\varepsilon}{2} \cos(kx + \theta) \end{aligned} \quad (7)$$

Let the phase screen have N photons per sub-aperture at the telescope pupil. Then the statistical fractional amplitude fluctuation is

$$\frac{RH - LH}{\sqrt{N}} = \frac{\sqrt{\frac{N}{4} + \delta N_{RH}} - \sqrt{\frac{N}{4} + \delta N_{LH}}}{\sqrt{N}} \quad (8)$$

where the incident phase screen amplitude is \sqrt{N} . A sinusoidal phase perturbation in radians creates a sinusoidal fractional amplitude of the same magnitude according to Eq (6)-(7). In the limit that N is large, we have $\delta N_{RH} \ll N$ and $\delta N_{LH} \ll N$ and:

$$\frac{RH - LH}{\sqrt{N}} \approx \frac{\delta N_{RH} - \delta N_{LH}}{N} \quad (9)$$

For Poisson statistics

$$\sqrt{\langle (\delta N_{RH})^2 \rangle} = \sqrt{\langle (\delta N_{LH})^2 \rangle} = \sqrt{\frac{N}{4}} \quad (10a)$$

The rms phase fluctuation in radians is the rms value of the quantity in Eq (9). Since $\langle \delta N_{RH} \delta N_{LH} \rangle = 0$, we have

$$\sigma = \sqrt{\frac{1}{2N}} \text{ radians} \quad (11a)$$

It is now simple to generalize for electron read noise too. Eq (10a) generalizes to

$$\langle (\delta N_{RH})^2 \rangle = \langle (\delta N_{LH})^2 \rangle = \frac{N}{4} + (\text{read } e^-)^2 \quad (10b)$$

Eq (11a) then generalizes to:

$$\sigma = \sqrt{\frac{1}{2N} + \frac{2(\text{read } e^-)^2}{N}} \text{ radians} \quad (11b)$$

Eq (11b) has been verified by simulations with a constant phase constant amplitude phase screen. The rms value of the reconstructed wavefront in the pupil closely matches prediction.

4. SIMULATIONS

The pyramid WFS simulator⁵ runs from a movie of phase screens generated using the ARROYO⁶ code. These phase screens were for AO simulations of the proposed thirty-meter telescope (TMT). They were generated on a 1025x1025 grid every 0.0005 seconds for a 4 second period. Distributed atmospheric turbulence typical at Cerro Pachon in the Chilean Andes ($r_0=0.15$ meters at $\lambda=500$ nm) was assumed. Each turbulence layer had a representative wind velocity.

Simulations were done for the pyramid pair WFS, the Mach-Zehnder WFS, and the Shack-Hartmann WFS. For all three kinds of WFS, the phase screen complex wavefronts were all spatially filtered keeping only those spatial frequencies having k_x and k_y values up to the Nyquist values. Physically, this corresponds to focusing through a square pinhole and then re-imaging. A 128x128 sub-aperture grid covered the pupil. The pyramid pair WFS simulations used a 144x144 sub-aperture grid for which the center 128x128 sub-aperture grid covered the pupil. The 1025x1025 phase screens were extended to 1152x1152. Spatial filtering reduced the size of the grids eight-fold to 144x144.

The simulations are for a pair of 2-sided pyramid wavefront sensors that have their respective ridges orthogonal to one another. This is because any speckle pair that falls along the ridge of a pyramid wavefront sensor will not be sensed by that wavefront sensor. It is therefore blind to phase variations that are purely along the ridge direction. Each separate wavefront sensor will be blind to either the pure x or the pure y phase variations. However, by using both together, all phase variations will be sensed.

The two orthogonal two-sided pyramid wavefront sensors are labelled RH-LH and TH-BH. The RH-LH sensor has right and left halves, whereas the TH-BH sensor has top and bottom halves. The RH-LH WFS measures the phase except for the y variation and the TH-BH WFS measures the phase except for the x variation. Their two measurements must be combined. The x variation is obtained from the RH-LH WFS and the y variation is obtained from the TH-BH WFS. The x and y variations are then removed from both and they are averaged together. The x and y variations are then restored.

The simulations were done with various frame rates for different magnitude stars. The simulations applied the corrections at the end of each frame. This is the limiting case of being able to compute and apply the corrections infinitely fast. In reality, the corrections will be applied sometime within the next frame. The control systems for both the optimistic limiting case and the general real case are discussed in Appendix A. We find that the system is stable for $0 < g < 2$ for the optimistic limiting case and is only stable for $0 < g < 1$ for the case in which the corrections are applied at the end of the next frame.

There is a double penalty to the effective bandwidth caused by any lag in applying the corrections. Let T be the integration time for each frame and let x be the lag in frames such that $x=0$ means that the corrections are applied immediately at the end of each frame and $x=1$ means that the corrections are applied at the end of the next frame for each frame. We look at each of these cases.

For the $x=0$ case, at the beginning of each frame just after corrections have been made, the deformable mirror (DM) is already $T/2$ out of date, because the corrections are based upon an average over the last frame of duration T . At the end of the frame just before the corrections are made, the DM is $3T/2$ out of date. The average that the DM is out of date is T . Since the control loop can operate at $g=1$, the effective bandwidth is T^{-1} .

For the case in which the corrections are made x frames later, we average how much out of date the DM is over a frame. We assume $x > 0$. Just after the corrections are made, the DM is $(x+1/2)T$ out of date. At the end of the frame the DM is $3T/2$ out of date. So at the beginning of a frame it is also $3T/2$ out of date. Consequently, just before the corrections are made at x , the DM is $(x+3/2)T$ out of date. The average that the DM is out of date is:

$$\frac{1}{2}x \left[\left(\frac{3}{2} \right) + \left(x + \frac{3}{2} \right) \right] T + \frac{1}{2}(1-x) \left[\left(x + \frac{1}{2} \right) + \left(\frac{3}{2} \right) \right] T = (1+x)T \quad (12)$$

For $x=1$, we would probably operate at $g=1/2$. This gives the same stability safety factor that we have operating at $g=1$ for $x=0$. The effective bandwidth is then $T^{-1}/4$. A lag of one frame has resulted in reducing the effective bandwidth by a factor of four, not two. This means that the wavefront errors that the pyramid wavefront sensor sees are four times as great. If the wavefront errors that the pyramid wavefront sensor sees are too large, it cannot operate in direct phase mode. What is important is what the errors are in radians. For any given x , T , and atmospheric conditions, the pyramid wavefront sensor can always operate in direct phase mode at a sufficiently long wavelength.

Figure 5 shows the result of a simulation at $t=0.100$ seconds for a 2000 Hz frame rate, 1014 photons/sub-aperture, and $\lambda_{WS}=0.85 \mu\text{m}$. The residual phase aberration in radians at the science camera wavelength is shown on the left while the modulus of the real FFT of the residual phase aberration is shown on the right. Piston was first subtracted from the phase residual before computing the FFT. Failure to subtract the piston would give rise to an additional Airy diffraction pattern which would fill in the center square dark region. This is roughly what the science camera amplitude would be with a coronagraph if scintillation were negligible. Figure 6 is a wide horizontal line-out of the square of the FFT amplitude shown in Figure 5b. The line-out is somewhat below center.

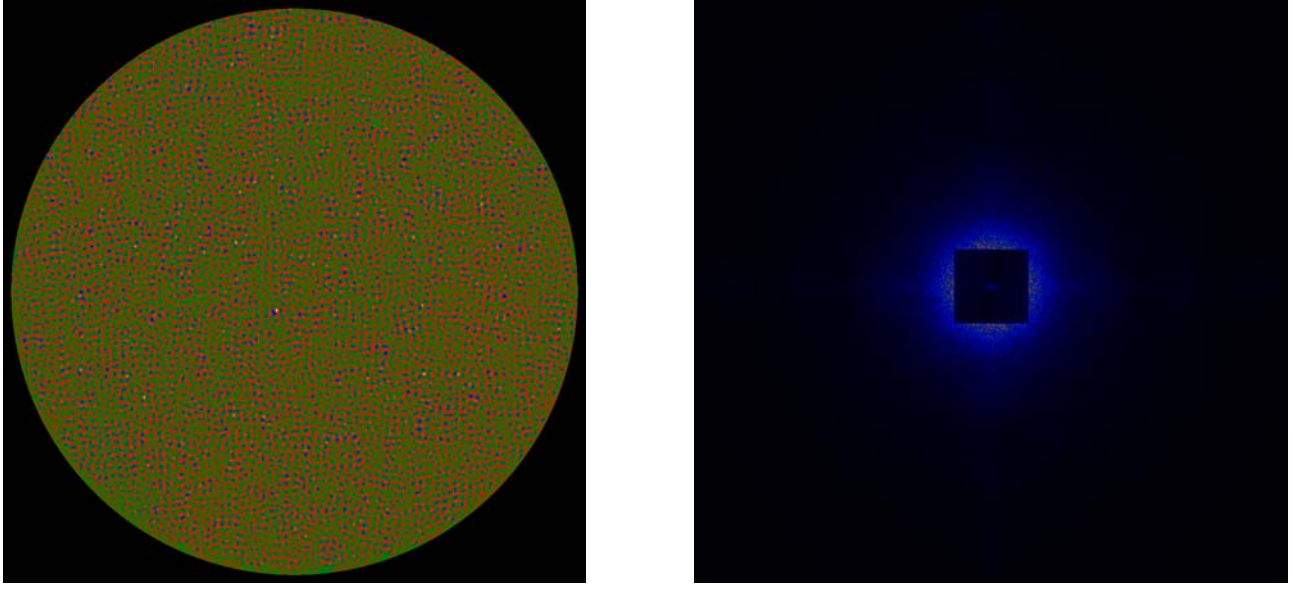


FIGURE 5 (a) residual phase aberration (b) modulus of its real FFT
Simulation for the pyramid WFS pair at $t=0.100$ seconds for 2000 Hz frame rate, 1014 photons/sub-aperture, $\lambda_{WS}=0.85 \mu\text{m}$

To estimate the science camera halo relative intensity without modelling the coronagraph, we first compute the modulus squared of the real FFT of the residual phase in radians at the science camera wavelength. The piston must be subtracted first and the FFT should be normalized so that the norm is preserved. The norm is the sum of the modulus squared values at all the pixels. Then compute the modulus squared of the FFT of the fractional amplitude variations. It is assumed that without scintillation the amplitude is a top hat. The image of the fractional amplitude variations should be zero outside the aperture. Now add the modulus-squared values of the two FFT's. The last step is to multiply by the inverse of the area fraction for the pupil. If the pupil is inscribed in the square grid, that factor is $4/\pi$.

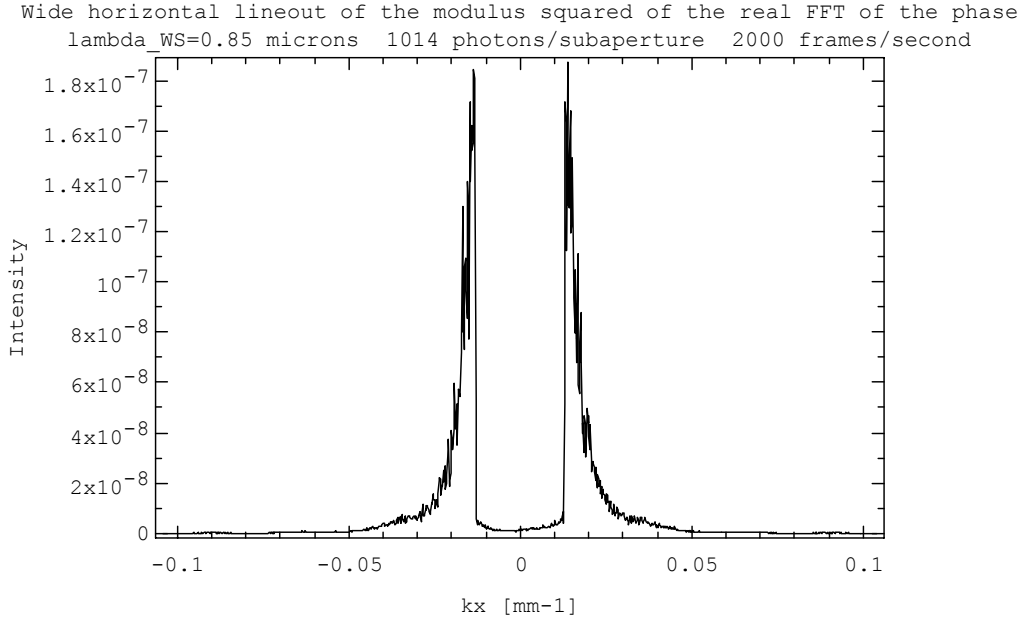


FIGURE 6
Modulus squared of the real FFT of the residual phase

The PSF was computed using Blackmann apodization and is shown in Figure 6. It includes the effect of the scintillation. It compares the 2-sided pyramid pair WFS with the Mach-Zehnder WFS and the Shack-Hartmann WFS. The 2029 and 225 photons per sub-aperture for the left and right figures represents 1014.5 and 112.5 photons per sub-aperture, respectively, for each of the two 2-sided pyramid wavefront sensors. For a bright star, all three wavefront sensors perform comparably. But

for a dim star, the 2-sided pyramid pair WFS performs best, and the Shack-Hartmann WFS performs worst. All three wavefront sensors are spatially filtered.

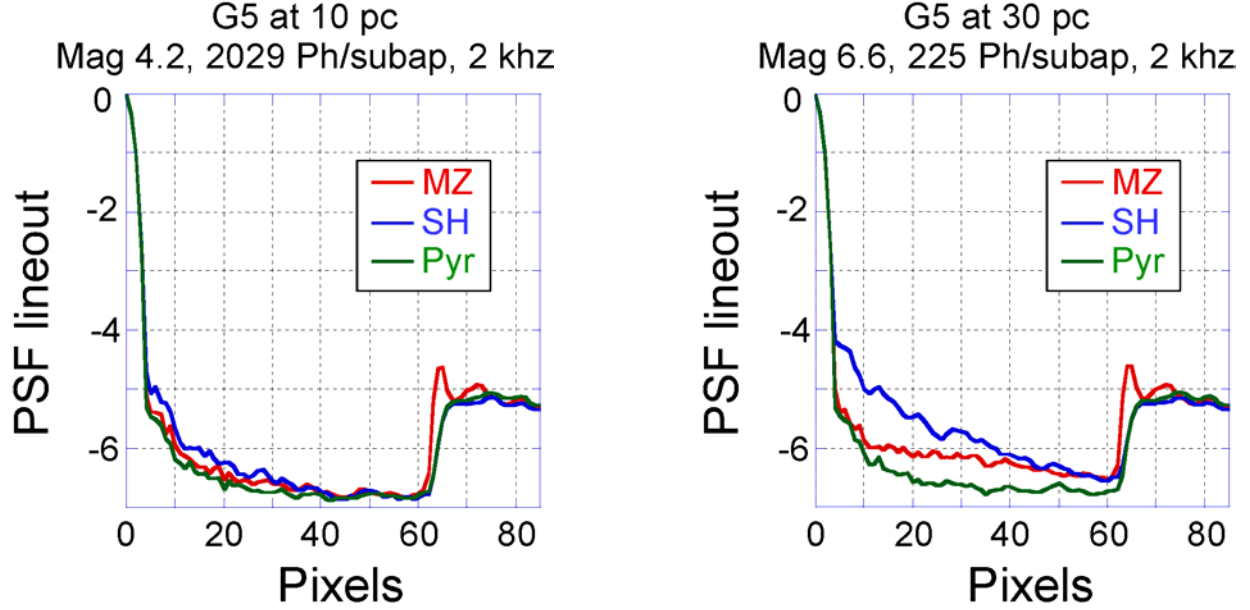


FIGURE 7 Comparison of the PSF for a Mach-Zehnder interferometer WFS (MZ), a Shack-Hartmann WFS (SH), and a two-sided pyramid pair WFS operating in the direct phase mode (Pyr) for a bright star and a dimmer star. All three wavefront sensors were spatially filtered. The 30 meter diameter telescope pupil was covered by a 128x128 sub-aperture array. The edge of the dark region is at the Nyquist spatial frequency.

5. COMPARISON WITH THREE AND FOUR SIDED PYRAMID WAVEFRONT SENSORS

The performance of the two-sided pyramid wavefront sensor has been compared with the performances of three- and four-sided pyramid wavefront sensors. The four-sided pyramid wavefront sensor has the problem that it is possible for speckle pairs to fall on one of the two orthogonal ridges. For the three-sided pyramid wavefront sensor, if one speckle falls on an edge, the other will not. Correction for the spatial frequencies must be done in both k_x and k_y .

Reconstruction for the four-sided pyramid wavefront sensor is still straightforward. We make the four-sided pyramid reconstruction problem into a two-sided pyramid reconstruction problem. We show how the RH-LH difference amplitude is computed. We begin by computing the real FFT's of the quadrant amplitudes. Label the quadrants Q1, Q2, Q3, and Q4 and let the quadrant amplitudes be A_{Q1} , A_{Q2} , A_{Q3} and A_{Q4} . Let the corresponding real FFT's of the quadrant amplitudes be \mathbb{A}_{Q1} , \mathbb{A}_{Q2} , \mathbb{A}_{Q3} and \mathbb{A}_{Q4} . Mask the FFT of each quadrant to its quadrant and the opposite quadrant. For instance, mask the FFT of the Q1 quadrant amplitude to quadrants Q1 and Q3. Denote these masked FFT's $\mathbb{A}_{Q1|Q1,Q3}$, $\mathbb{A}_{Q2|Q2,Q4}$, $\mathbb{A}_{Q3|Q3,Q1}$, and $\mathbb{A}_{Q4|Q4,Q2}$. The core is divided up. The edges must be treated specially, but we will discuss that later. Compute the complex masked differences:

$$\mathbb{A}_{Q1|Q1,Q3} - \mathbb{A}_{Q3|Q3,Q1} \quad (13a)$$

$$\mathbb{A}_{Q4|Q4,Q2} - \mathbb{A}_{Q2|Q2,Q4} \quad (13b)$$

Now make the spatial frequency corrections in k_x and k_y for each of these complex differences. The reason that is possible to do the spatial frequency corrections at this point is that the core is subtracted out in these two differences and the centroid shifts of the core are equal and opposite in each difference. Then take the real part of the inverse FFT of the sum of $\mathbb{A}_{Q1|Q1,Q3} - \mathbb{A}_{Q3|Q3,Q1}$ and $\mathbb{A}_{Q4|Q4,Q2} - \mathbb{A}_{Q2|Q2,Q4}$. This is the RH-LH difference amplitude. We have made the four-sided pyramid reconstruction problem into a two-sided pyramid reconstruction problem. In the masking, speckles on the +y axis are made to contribute to the RH side and speckles on the -y axis are made to contribute to the LH side.

Reconstruction for the three-sided pyramid wavefront sensor is complicated. We again make the pyramid reconstruction problem into a two-sided pyramid reconstruction problem by computing the RH-LH difference amplitude. We begin by computing the real FFT's of the sector amplitudes. Now adding pieces of FFT's is much more complicated. Let the sectors be $S_{0^\circ-120^\circ}$, $S_{120^\circ-240^\circ}$, and $S_{240^\circ-360^\circ}$. Angles are measured counterclockwise from the +x axis towards the +y axis. Let $\mathbb{S}_{0^\circ-120^\circ}$, $\mathbb{S}_{120^\circ-240^\circ}$, and $\mathbb{S}_{240^\circ-360^\circ}$.

S_{240° , and $S_{240^\circ-360^\circ}$ be the corresponding real FFT's of the amplitudes. Let $S_{0^\circ-120^\circ | 0^\circ-60^\circ, 180^\circ-240^\circ}$ be $S_{0^\circ-120^\circ}$ masked to include only 0° to 60° and 180° to 240° . Continue to use this notation. Construct the complex masked differences:

$$S_{0^\circ-120^\circ | 0^\circ-60^\circ, 180^\circ-240^\circ} - S_{120^\circ-240^\circ | 180^\circ-240^\circ, 0^\circ-60^\circ} \quad (14a)$$

$$S_{0^\circ-120^\circ | 60^\circ-90^\circ, 240^\circ-270^\circ} - S_{240^\circ-360^\circ | 240^\circ-270^\circ, 60^\circ-90^\circ} \quad (14b)$$

$$S_{240^\circ-360^\circ | 300^\circ-360^\circ, 120^\circ-180^\circ} - S_{120^\circ-240^\circ | 120^\circ-180^\circ, 300^\circ-360^\circ} \quad (14c)$$

$$S_{240^\circ-360^\circ | 270^\circ-300^\circ, 90^\circ-120^\circ} - S_{0^\circ-120^\circ | 90^\circ-120^\circ, 270^\circ-300^\circ} \quad (14d)$$

It is at this point (Eq 13) that the spatial frequencies were corrected for the 4-sided pyramid WFS. However, it is not possible to correct the spatial frequencies for a pyramid WFS with an odd number of sectors. This is because the core centroids in the two sectors from which the FFT pieces are taken in each difference do not have opposite shifts. The real part of the inverse FFT of the sum of the four differences in Eq (14) gives the RH-LH amplitude difference. The three-sided pyramid reconstruction problem has been made into a two-sided pyramid reconstruction problem. The edges require special treatment. The time-consuming reconstruction makes the three-sided pyramid wavefront sensor impractical for an AO system with a high frame rate.

The 2-sided, 3-sided, and 4-sided pyramid wavefront sensors were compared by doing six iterations with a static phase screen having $D=0.25$, $W=1$, and $r_0=2$. These simulations were done using the *Telescope AO code*⁵ and are summarized in Table 2.

Iteration	2-sided [radians] (k_x corrected)	3-sided [radians] (k_x, k_y not corrected)	4-sided [radians] (k_x, k_y not corrected)	4-sided [radians] (k_x, k_y corrected)
1	0.0088	0.0528	0.0238	0.0123
2	0.0017	0.0219	0.0119	0.0044
3	0.0006	0.0401	0.0082	0.0019
4	0.0005	0.0269	0.0063	0.0010
5	0.0005	0.0312	0.0050	0.0007
6	0.0005	0.0281	0.0046	0.0007

TABLE 2 The 2-sided pyramid pair WFS with k_x corrected performs best. The 4-sided pyramid WFS with k_x and k_y corrected performs acceptably. The 3-sided pyramid WFS performs unacceptably poorly. Its errors were low order in the interior.

6. SUMMARY

The 2-sided pyramid pair wavefront sensor performs much better in the direct phase mode than does the 4-sided pyramid wavefront sensor. Its noise properties are the same. However, it will require additional optics. For the 2-sided pyramid wavefront sensor, the reconstruction obtained from the amplitude difference of the two sides is much more accurate than the reconstruction obtained from the intensity difference of the two sides. Correction of the spatial frequencies greatly improves the performances of both the 2-sided and the 4-sided pyramid wavefront sensors in the direct phase mode. The three-sided pyramid sensor performs poorly. Correction of the spatial frequencies cannot be done for any pyramid sensor with an odd number of sectors.

ACKNOWLEDGEMENTS

This research was performed under the auspices of the U.S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under Contract W-7405-ENG-48, and also supported in part by the National Science Foundation Science and Technology Center for Adaptive Optics, managed by the University of California at Santa Cruz under cooperative agreement No. AST - 9876783.

REFERENCES

1. Robert Ragazzoni, "Pupil plane wave sensing with an oscillating prism", *J. Mod. Opt.* **43** No. 2, 1996. pp 289-293
2. C. Verinaud, "On the nature of the measurements provided by a pyramid wave-front sensor", *Opt. Comm.* 233, 2004, pp 27-38
3. Olivier Guyon, "Limits of Adaptive Optics for high contrast imaging", *Astrophys. J.* 629 (2005) pp 592-614
4. C. Verinaud, M. Le Louarn, V. Korkiakoski, and M. Carbillet, "Adaptive Optics for high-contrast imaging pyramid sensor versus spatially filtered Shack-Hartmann sensor", *Mon. Not. R. Astron. Soc.* 357, L26-L30, 2005.

5. The *Telescope AO* code is written in Microsoft Visual C++ and requires Windows XP Professional. It has the pyramid WFS simulator and can be obtained from phillion1@llnl.gov
6. Matthew C. Britton, “Arroyo”, Proc. of SPIE Vol. 5497 (SPIE, Bellingham, WA, 2004), pp 290-300

APPENDIX A MAXIMUM GAIN FOR CLOSED LOOP STABILITY

The maximum gain for closed loop stability can be determined using a static phase screen. Let DM_N be the deformable mirror wavefront after the N^{th} correction and let PS be the static phase screen wavefront. Since it is static, it has no subscript. For each frame, let the correction be made at x , where $x=0$ means that the correction is made immediately after the frame, and $x=1$ means that the correction isn't made until the end of the next frame. The quantity x represents the lag in frames. The control law is then

$$DM_N = DM_{N-1} + g \left[x(PS - DM_{N-2}) + (1-x)(PS - DM_{N-1}) \right] \quad (\text{A1})$$

Here g is the gain. If $g=1$, then the deformable mirror (DM) is incremented by the measured wavefront residual aberration, which we assume to be the actual wavefront residual aberration.

It may be disconcerting to see DM_N and PS being treated as scalar quantities even though they are actually two-dimensional arrays. Just think of Eq (A1) as being for a particular mode and DM_N and PS as being coefficients of that particular modal function. Eq (A1) is true for any modal function.

In order to study the stability, look at the homogeneous difference equation obtained by setting PS=0. It has solutions of the form $DM_N = \alpha r^N$. This gives a quadratic equation in r . The system is stable if and only if both roots have modulus less than one. The roots are given by:

$$r_{\pm} = \frac{1}{2} \left[1 - g + g x \pm \sqrt{(1 - g + g x)^2 - 4 g x} \right] \quad (\text{A2})$$

For $x=0$, the roots are 0 and $1-g$. The system is stable if $0 < g < 2$. For $x=1$, the roots are

$$r_{\pm} = \frac{1}{2} \left(1 \pm \sqrt{1 - 4g} \right) \quad (\text{A3})$$

For $0 < g < 1/4$, both roots are real and have modulus less than one. For $g > 1/4$, both roots have modulus \sqrt{g} . Therefore, the system is stable if and only if $0 < g < 1$.